Modeling Interventions to Improve Access to Public Health Information

Elizabeth D. Liddy, PhD, Anne M. Turner, MD, Jana Bradley, PhD
Center for Natural Language Processing, Syracuse University, Syracuse, NY and Oregon Health & Science University, Portland, OR

Abstract: In a Robert Wood Johnson funded project, we are establishing a model-based means for automatically analyzing and representing grey literature that reports on public health (PH) interventions We summarize the development of an intervention model for public health documents and provide a project update on the implementation of natural language technology to improve access to difficult to find public health information.

Introduction: Reports concerning public health interventions are frequently not available through commercial publishers. These grey literature reports are increasingly placed on the Internet but are poorly organized and difficult to access. Efficiency of access and organization of this literature is critical to enable PH professionals to learn from and build on the activities of others and to provide assessments and examples of successful interventions for stakeholders.

Methods and Results: We collected a representative sample of 300 grey literature documents from 4 websites that contained a large number of public health documents: the New York Academy of Medicine's Grey Literature Report and the websites for health departments of Minnesota, Hennepin County, and Lake County. Documents were selected if they dealt with some aspect of a public health intervention and they were not from commercial publisher. We developed a preliminary model of the key elements of PH interventions based on an analysis of these 300 documents.

Having developed this model data-up, our next goal was to identify which components PH professionals believed were important for inclusion in a summary representation of a PH report. We validated our model of key elements using the input of PH professionals recruited from 4 professional PH listservs. The initial sample of 23 participants was selected based on the order of their response and their professional background and position. The goal was to have the sample reflect the make up of the PH workforce. Each participant was sent 4 sample documents from the larger collection to read and identify the key components of information. They were also asked to write an abstract for each document of the length and content they would want in order to determine if a document was useful. This input was used to modify the data-up set of components and prioritize them for automatic extraction. Some highlights of the results include: 1) all of the subjects' abstracts included a problem statement, a description of the intervention or purpose of the report; 2) publication date or research date, as well as the group responsible for creating the document were frequently noted, and; 3) abstracts written for demographic reports included parameters studied and a summary of the results.

Based on an analysis of the documents and the input of the PH professionals, we produced a model of the important elements of the PH intervention reporting. This specialist-validated model was used to guide the development of Natural Language Processing (NLP) algorithms which process full text PH reports to identify the key model components that the system should extract in order to produce a rich, yet concise representation of each PH report. We will next focus on developing access and interface technologies that will best utilize these brief, but rich, model-based representations of PH grey literature reports.

Conclusions: Although the PH grey literature is diffuse in subject and format, a review of 300 PH grey literature documents revealed that the literature can be represented by a single PH intervention model. Furthermore, input from PH professionals substantiated our initial intervention model. Key elements include identification and description of a PH problem, the individuals or groups addressing the problem, the target population, the proposed intervention, and evaluation of the intervention's outcomes. We have shown that lengthy reports can be successfully trimmed to enable processing of those document sections which provide rich areas for automatically detecting and extracting PH intervention elements. The goal of producing technology for automatically extracting key elements of PH grey literature reports is to provide important information to PH professionals in deciding if a particular document will be useful to their current situation, and this goal is in sight.

Acknowledgments: This work was funded by the Robert Wood Johnson Foundation (Grant # 04668).